

# Branching Process approach for 2-SAT thresholds

Elchanan Mossel\*      Arnab Sen †

May 20, 2009

## Abstract

It is well known that, as  $n$  tends to infinity, the probability of satisfiability for a random 2-SAT formula on  $n$  variables, where each clause occurs independently with probability  $\alpha/2n$ , exhibits a sharp threshold at  $\alpha = 1$ . We study a more general 2-SAT model in which each clause occurs independently but with probability  $\alpha_i/2n$  where  $i \in \{0, 1, 2\}$  is the number of positive literals in that clause. We generalize branching process arguments by Verhoeffen(99) to determine the satisfiability threshold for this model in terms of the maximum eigenvalue of the branching matrix.

**Keywords:** 2-SAT, satisfiability, phase transition, 2-type branching process.

---

\*Dept. of Statistics, U.C. Berkeley. Supported by an Alfred Sloan fellowship in Mathematics, NSF grants DMS 0528488 and DMS 0548249 (Career), and ONR grant N0014-07-1-05-06.

†Dept. of Statistics, U.C. Berkeley. Supported by NSF grants DMS 0528488 and DMS 0548249 and ONR grant N0014-07-1-05-06.

# 1 Introduction

## 1.1 Background

The  $k$ -satisfiability (in short  $k$ -SAT) problem is a canonical constraint satisfaction problem in theoretical computer science. A  $k$ -SAT formula is a conjunction of  $m$  clauses, each of which is a disjunction of length  $k$  chosen from  $n$  boolean variables and their negations. Given a  $k$ -SAT formula a natural question is to find an assignment of  $n$  variables which satisfies the formula. The decision version of the problem is to determine whether there exists an assignment satisfying the formula.

From the computational complexity perspective, the problem is well understood. The problem is NP-hard for  $k \geq 3$  [Coo71] and linear time solvable for  $k = 2$  [APT82]. Much recent interest was devoted to the understanding of random  $k$ -SAT formulas where each clause is chosen independently with the same probability and the expected number of clauses in the formula is  $\alpha n$ . This problem lies in the intersection of three different subjects — statistical physics, discrete mathematics and complexity theory.

In statistical physics, the notion of ‘phase transition’ refers to a situation where systems undergo some abrupt behavioral change depending on some external control parameter such as temperature. In the context of random  $k$ -SAT formulas the natural parameter is the density of the formula  $\alpha$ , i.e., the ratio between the number of clauses to the number of variables. Much recent research is devoted to understanding the critical densities for random  $k$ -SAT problems. The most important one is the critical density for satisfiability, i.e., the threshold at which a formula becomes from satisfiable with high probability to unsatisfiable with high probability [MMZ06, AP04]. Other thresholds involve the geometry of the solution space and the performance of various algorithms (e.g. see [ART06, MMW07] and references therein).

The problem of 2-SAT is more amenable to analysis than  $k$ -SAT for  $k \geq 3$ . This is closely related to the fact that 2-SAT can be solved in linear time and to a clear graph theoretic criteria which is equivalent to satisfiability. The threshold for 2-SAT is well known to be  $\alpha = 1$  ( see [Goe92, Goe96, CR92, FdI92] ) and detailed information on the scaling window is given in [BBC<sup>+</sup>01]. For  $k$ -SAT for  $k \geq 3$  there are various bounds and conjectures on the critical threshold for satisfiability but the thresholds are not known rigorously [Fri99, AP04, Ach00, PZ02].

In the current paper we establish the threshold of a more general 2-SAT model where the probability of having a clause in the formula depends on the number of positive and negative variables in the clause. Our proof is based on branching process arguments. Branching process techniques have been used before to study the standard 2-SAT formulas in the unsatisfiability regime ( $\alpha > 1$ ), for example see [Ver99]. We generalize the arguments given in [Ver99] in the 2-type branching process set-up to analyze the general 2-SAT model. Our main contribution is in demonstrating that branching process arguments extend to a multi-type setup. A well accepted idea in studying random graphs and constraint satisfaction problems is that since “local” structure of the problems is tree-like, processes defined on trees play a key role in analyzing the problems. The classical example is the threshold for the existence of a “giant” component in random graphs where branching processes play a key role in the proof (see e.g [JLR00, Bol01]). Some more recent examples include [Ald01, Wei06, MWW08].

A seemingly closely related work is by Cooper et. al. [CFS02] where the threshold for random 2-SAT with given literal degree distribution is derived. We note that the two papers are incomparable, since ours is stated in terms of the distribution of clauses of different types while in [CFS02] the model is stated in terms of the degrees of the literals. While the distribution of clauses determines the distribution of the degrees of the literals, the converse does not hold. For example, a random 2-SAT formula with  $2n$  positive-negative clauses has the same literal degree distribution as a uniform random formula with  $2n$  clauses. It is obvious that while the former is always satisfiable, the latter is not satisfiable with high probability.

## 1.2 Definitions and statements of main results

Let  $x_1, x_2, \dots, x_n$  be  $n$  boolean *variables*. Calling the negation of  $x_i$  as  $\bar{x}_i$ , these  $n$  boolean variables give us  $2n$  *literals*  $\{x_1, x_2, \dots, x_n, \bar{x}_1, \bar{x}_2, \dots, \bar{x}_n\}$ . The two literals  $x_i$  and  $\bar{x}_i$  are called complementary to each other ( $x_i = 1$  iff  $\bar{x}_i = 0$ ) with the convention that  $\bar{\bar{x}}_i = x_i$ . We will call the  $n$  literals  $x_i, i = 1, 2, \dots, n$  *positive literals* and their complementary literals  $\bar{x}_i, i = 1, 2, \dots, n$  *negative literals*.

Given a literal  $u$ ,  $\text{var}(u)$  denotes the corresponding variable, the notation naturally extending to a set of literals  $S$  by  $\text{var}(S) = \{\text{var}(u) : u \in S\}$ . Two literals  $u$  and  $v$  are said to be *strongly distinct* if  $u \neq v$  and  $u \neq \bar{v}$ , equivalently, if  $\text{var}(u) \neq \text{var}(v)$ . A *2-clause* (which we will call simply a “clause” later) is a disjunction  $C = x \vee y$  of two strongly distinct literals. In this paper we will not allow  $x \vee x$  or  $x \vee \bar{x}$  as valid clauses. A 2-SAT formula is a conjunction  $F = C_1 \wedge C_2 \wedge \dots \wedge C_m$  of 2-clauses  $C_1, C_2, \dots, C_m$ . Let  $\mathcal{C} = \{C_1, C_2, \dots, C_m\}$  be the collection of clauses corresponding to  $F$ .

As usual in the boolean algebra, 0 stands for the logical value FALSE, and 1 stands for the logical value TRUE. A 2-SAT formula  $F = F(x_1, x_2, \dots, x_n)$  is said to be *satisfiable* if there exists a truth assignment  $\eta = (\eta_1, \eta_2, \dots, \eta_n) \in \{0, 1\}^n$  such that  $F(\eta_1, \eta_2, \dots, \eta_n) = 1$ . The formula  $F$  is called *SAT* if the formula  $F$  is satisfiable and  $F$  is called *UNSAT* otherwise.

In the standard model for a random 2-SAT formula we choose each of the possible  $4 \binom{n}{2}$  clauses independently with probability  $\alpha/2n$ . In this paper we will study a more general model in which a 2-SAT formula  $F$  consists of a random subset  $\mathcal{C}$  of clauses such that each clause appears in  $\mathcal{C}$  independently and a clause having  $i$  positive literals is present in the formula with probability  $\alpha_i/2n, i = 0, 1, 2$  for some constants  $\alpha_i \geq 0$ . Of course, taking  $\alpha_0 = \alpha_1 = \alpha_2 = \alpha$  we retrieve the standard model.

Let  $\mathbf{M}$  be the branching matrix given by

$$\mathbf{M} = \frac{1}{2} \begin{bmatrix} \alpha_1 & \alpha_0 \\ \alpha_2 & \alpha_1 \end{bmatrix}. \quad (1)$$

Note that though  $\mathbf{M}$  is not symmetric in general, its eigenvalues are all real and given by  $\frac{1}{2}(\alpha_1 \pm \sqrt{\alpha_0\alpha_2})$ . Let  $\rho = \frac{1}{2}(\alpha_1 + \sqrt{\alpha_0\alpha_2})$  denote the largest eigenvalue of  $M$ . We show that  $\rho$  is a crucial parameter for satisfiability. In particular, our main result is the following theorem which establishes that the generalized 2-SAT model undergoes a phase transition from satisfiability to unsatisfiability at  $\rho = 1$ .

**Theorem 1** *Let  $F$  be random 2-SAT formula under generalized model with parameter  $\rho$ .*

- (a) *If  $\rho < 1$  or  $\alpha_0\alpha_2 = 0$  then  $F$  is satisfiable with probability tending to one as  $n \rightarrow \infty$ .*
- (b) *If  $\rho > 1$  and  $\alpha_0\alpha_2 > 0$  then  $F$  is unsatisfiable with probability tending to one as  $n \rightarrow \infty$ .*

**Remark 1** *It is easy to see and well known that the satisfiability threshold for 2-SAT remains the same for variants of the model where the set of 2-clauses contains also clauses of the form  $x \vee y$  where  $x$  and  $y$  may not be strongly distinct. Similarly, the threshold remains the same if  $n\alpha$  clauses are chosen uniformly at random instead of choosing each clause independently with probability  $\alpha/2n$  (See the appendix A of [BBC<sup>+</sup>01]). The same reasoning applies to the more general 2-SAT model considered here.*

## 2 2-SAT and the implication digraph

We will exploit the standard representation of a 2-SAT formula as a directed graph (see [BBC<sup>+</sup>01] for example), called the *implication digraph* associated with the 2-SAT formula. This graph has  $2n$  vertices, labelled by the  $2n$  literals. If the clause  $(u \vee v)$  is present in the 2-SAT formula then we draw the two directed edges  $\bar{u} \rightarrow v$  and  $\bar{v} \rightarrow u$ . The directed edges can be thought of as logical implications since if there is a directed edge from  $u \rightarrow v$  and  $u = 1$ , then for the formula to be satisfiable, it is necessary to have  $v = 1$ .

By a *directed path* (from  $u$  to  $v$ ), we mean a sequence of vertices  $u_0 = u, u_1, u_2, \dots, u_k = v$  such that there is a directed edge from  $u_i \rightarrow u_{i+1}$  for all  $i = 0, 1, \dots, k - 1$ . The length of this directed path is  $k$ . A *contradictory cycle* is a union of two (not necessarily disjoint) directed paths - one starts from a literal  $u$  and ends at its compliment  $\bar{u}$ , the other starts from  $\bar{u}$  and ends at  $u$ .

The following lemma connects the concept of satisfiability of the 2-SAT problem to the existence of contradictory cycle in the implication digraph. For a proof, see [BBC<sup>+</sup>01].

**Lemma 1** *A 2-SAT formula is satisfiable iff its implication digraph contains no contradictory cycle.*

### 3 Proof of Theorem 1 part (a)

The proof has some resemblance with the first moment arguments given in Chvátal and Reed [CR92]. In fact our ‘hooked chain’ is same as what they called ‘bicycle’. The extension to the more general case considered here uses a recursive argument which allows us to deal with the multi-parameter general model.

**Definition 1** *Suppose there exists strongly distinct literals  $y_1, y_2, \dots, y_s$  and  $u, v \in \{y_1, y_2, \dots, y_s, \bar{y}_1, \bar{y}_2, \dots, \bar{y}_s\}$  such that  $(\bar{u} \vee y_1), (\bar{y}_1 \vee y_2), \dots, (\bar{y}_{s-1} \vee y_s), (\bar{y}_s \vee v) \in \mathcal{C}$  or equivalently, there exists  $u \rightarrow y_1 \rightarrow y_2 \rightarrow \dots \rightarrow y_s \rightarrow v$  in the implication digraph corresponding to the 2-SAT formula. We call this sequence of literals a hooked chain (of length  $s + 1$ ).*

**Lemma 2** *If a 2-SAT formula is unsatisfiable then its implication digraph contains a hooked chain of length  $\geq 3$ .*

**Proof.** Suppose a 2-SAT formula is unsatisfiable. By lemma 1, we have a contradictory cycle in the implication digraph, say  $u_0 \rightarrow u_1 \rightarrow u_2 \rightarrow \dots \rightarrow u_l = \bar{u}_0 \rightarrow u_{l+1} \rightarrow u_{l+2} \rightarrow \dots \rightarrow u_k = u_0$ . The cycle has at least one directed path from a literal to its complement. Choosing one that minimizes the length we get an implication chain formed by a sequence of the literals  $u_h \rightarrow u_{(h+1) \bmod k} \rightarrow u_{(h+2) \bmod k} \rightarrow \dots \rightarrow u_{(h+t) \bmod k} = \bar{u}_h$  so that  $u_{(h+1) \bmod k}, u_{(h+2) \bmod k}, \dots, u_{(h+t) \bmod k}$  are strongly distinct. Find the largest  $s \geq t$  such that  $u_{(h+1) \bmod k}, u_{(h+2) \bmod k}, \dots, u_{(h+s) \bmod k}$  are strongly distinct. Let  $v$  be the element pointed to by  $u_{(h+s) \bmod k}$  in the cycle. Then clearly  $u_h \rightarrow u_{(h+1) \bmod k} \rightarrow u_{(h+2) \bmod k} \rightarrow \dots \rightarrow u_{(h+s) \bmod k} \rightarrow v$  is a hooked chain of length  $s + 1$ . Since there can be no edge between a literal  $w$  and its complement  $\bar{w}$ , we must have  $t \geq 2$  and therefore,  $s \geq 2$ .  $\square$

**Lemma 3**

$$\mathbb{P}(F \text{ is unsatisfiable}) \leq C \sum_{s=2}^n \frac{(2s)^2}{n} [T_{s-1}^+ + T_{s-1}^-] \quad (2)$$

where  $T_{s-1}^+$  (or  $T_{s-1}^-$ ) is the expected number of directed paths of length  $(s - 1)$  started from  $x_1$  (or  $\bar{x}_1$ ) consisting of strongly distinct literals with  $T_0^+ = T_0^- = 1$  and  $C = [\max(\alpha_0, \alpha_1, \alpha_2)]^2$ .

**Proof.** Let  $H_s$  be the number of hooked chain of length  $(s + 1)$  in the implication digraph for 2-SAT and  $\Gamma_s$  be the number of directed paths of  $s$  strongly distinct literals in the same digraph. From Lemma 2,

$$\mathbb{P}(F \text{ is unsatisfiable}) \leq \mathbb{P}(\exists \text{ a hooked chain of length } (s + 1) \text{ for some } s \geq 2) \quad (3)$$

$$\leq \sum_{s=2}^n \mathbb{E}(H_s) \quad (4)$$

$$\leq C \sum_{s=2}^n \frac{(2s)^2}{(2n)^2} \mathbb{E}(\Gamma_s) \quad (5)$$

$$= C \sum_{s=2}^n \frac{s^2}{n} [T_{s-1}^+ + T_{s-1}^-] \quad (6)$$

Step (4) follows from simple union bound and Markov inequality. Inequality (5) can be explained as following:

$$H_s = \sum_{y_1, \dots, y_s}^* \sum_{u, v \in \{y_1, \dots, y_s, \bar{y}_1, \dots, \bar{y}_s\}} \mathbb{I}((\bar{u} \vee y_1), (\bar{y}_1 \vee y_2), \dots, (\bar{y}_s \vee v) \in \mathcal{C}).$$

Here  $\sum^*$  means the sum is taken over all possible set of strongly distinct literals of size  $s$ . Observe that for a fixed choice of strongly distinct literals there are  $2s$  choices for each  $u$  and  $v$  and each clause occurs with probability at most  $\max(\alpha_0, \alpha_1, \alpha_2)/2n$ . Now taking expectation and using independence between clauses, we have

$$\begin{aligned} \mathbb{E}(H_s) &= \sum_{y_1, \dots, y_s}^* \sum_{u, v \in \{y_1, \dots, y_s, \bar{y}_1, \dots, \bar{y}_s\}} \mathbb{P}((\bar{y}_1 \vee y_2), \dots, (\bar{y}_{s-1} \vee y_s) \in \mathcal{C}) \mathbb{P}((\bar{u} \vee y_1) \in \mathcal{C}) \mathbb{P}((\bar{y}_s \vee v) \in \mathcal{C}) \\ &\leq \frac{C(2s)^2}{(2n)^2} \sum_{y_1, \dots, y_s}^* \mathbb{P}((\bar{y}_1 \vee y_2), (\bar{y}_2 \vee y_3), \dots, (\bar{y}_{s-1} \vee y_s) \in \mathcal{C}) \\ &= \frac{C(2s)^2}{(2n)^2} \mathbb{E}(\Gamma_s). \end{aligned}$$

Noting that the quantities  $T_{s-1}^+$  or  $T_{s-1}^-$  do not depend on  $x_1$ , step (6) follows.  $\square$

**Lemma 4** Write  $\mathbf{T}_k = (T_k^+, T_k^-)^T$ . Then

$$\mathbf{T}_{s-1} \leq \mathbf{M}^{s-1} \mathbb{1} \quad (7)$$

where  $\mathbf{M}$  is defined in (1) and  $\mathbb{1} = (1, 1)^T$ .

**Proof.** For a literal  $u$  strongly distinct from  $x_1$ , let  $J_u$  denote the number of directed paths of length  $(s-2)$  starting from  $u$  consisting of strongly distinct literals and not involving the variable  $x_1$ . Then

$$\begin{aligned} T_{s-1}^+ &= \sum_{\substack{u: u \text{ literals} \\ \text{var}(u) \neq \text{var}(x_1)}} \mathbb{E}[J_u \times \mathbb{I}((\bar{x}_1 \vee u) \in \mathcal{C})] \\ &= \sum_{\substack{u: u \text{ literals} \\ \text{var}(u) \neq \text{var}(x_1)}} \mathbb{E}[J_u] \times \mathbb{P}((\bar{x}_1 \vee u) \in \mathcal{C}). \end{aligned}$$

The last step follows from the independence of clauses.

Simple coupling argument yields  $\mathbb{E}[J_u] \leq T_{s-2}^+$  or  $T_{s-2}^-$  depending whether the literal  $u$  is positive or negative. Combining the above facts, we get the following recursive inequality,

$$\begin{aligned} T_{s-1}^+ &\leq n T_{s-2}^+ \mathbb{P}(\bar{x}_1 \vee x_2) + n T_{s-2}^- \mathbb{P}(\bar{x}_1 \vee \bar{x}_2) \\ &\leq \frac{\alpha_1}{2} T_{s-2}^+ + \frac{\alpha_0}{2} T_{s-2}^- \end{aligned} \quad (8)$$

and similarly,

$$T_{s-1}^- \leq \frac{\alpha_2}{2} T_{s-2}^+ + \frac{\alpha_1}{2} T_{s-2}^- \quad (9)$$

Now the above two equations can be written in a more compact way as follows

$$\mathbf{T}_{s-1} \leq \mathbf{M} \mathbf{T}_{s-2} \quad (10)$$

Iterating (10), we get  $\mathbf{T}_{s-1} \leq \mathbf{M}^{s-1} \mathbf{T}_0 = \mathbf{M}^{s-1} \mathbb{1}$ .  $\square$

**Proof of theorem 1 part (a).** We are now ready to finish the proof of part (a) of Theorem 1. If  $\alpha_0 \alpha_2 = 0$  then either all zero or all one assignment always satisfies it. So, take  $\alpha_0 \alpha_2 > 0$ . Then  $\mathbf{M}$  is semisimple (i.e. similar to a diagonal matrix).

By lemma 4,  $T_{s-1,n}^+ + T_{s-1,n}^- \leq \mathbb{1}^T \mathbf{M}^{s-1} \mathbb{1} \leq B \rho^{s-1}$  for some constant  $B$ . The last inequality holds since we assume  $\mathbf{M}$  is semisimple. Plugging it in (2), we finally have

$$\begin{aligned} \mathbb{P}(F \text{ is unsatisfiable}) &\leq \frac{K}{n} \sum_{s=2}^n s^2 \rho^{s-1} \text{ for some constant } K > 0 \\ &\leq O(n^{-1}) \quad \text{since } \rho < 1. \end{aligned} \quad (11)$$

$\square$

## 4 The Exploration Process

Observe that when  $\rho > 1$ , we need to find a contradictory cycle in the implication digraph of the random 2-SAT formula with high probability. In order to prove this, we will show that starting from any fixed vertex there is a constant probability that it *implies* a large number of literals in the digraph, meaning that there are directed paths to a large number of vertices from the fixed vertex. To achieve this, we explore the digraph dynamically starting from a fixed literal  $x$  under certain rules and keep track of variables that are implied by  $x$  at each step. We call this the *exploration process* which is defined next.

**Definition and Notations.** Given a realization of the 2-SAT formula and an arbitrarily fixed literal  $x$ , we will consider an exploration process in its implication digraph starting from  $x$ .

- This process describes the evolution of two sets of literals which will be called the *exposed set* and the *active set*.
- A literal is said to be *alive* in a particular step of the process if it is strongly distinct from those in the exposed set and from those in the active set at that step.
- We maintain two stacks for the literals in active set, one for positive literals and another for negative ones.
- At each step we pop-up a literal (call it the *current literal*) from one of the two stacks of active set, depending on some event to be described later and *expose* it. It means that we look for all the literals that are alive at that time and to which there is a directed edge from the current literal.
- We then put those new literals in the stacks of the active set (positive or negative) in some predetermined order and throw the current literal in the exposed set.
- We go on repeating this procedure until the stack of the active literals becomes empty and the process stops.

- Mathematically, let  $E_t$  and  $A_t$  denote respectively the set of exposed and active set of literals at time  $t$ . Also let  $U_t = \{var(u) : var(u) \notin var(E_t) \cup var(A_t)\}$  be the set of alive variables at time  $t$ . Set  $E_0 = \emptyset, A_0 = \{x\}$ . If  $A_t$  is non-empty and the literal  $l \in A_t$  is exposed from the stack, then we have the following updates at time  $t + 1$ ,

$$A_{t+1} = (A_t \setminus \{l\}) \cup \{u : u \text{ literal s.t. } var(u) \in U_t \text{ and the clause } (\bar{l} \vee u) \text{ is present}\}, \quad E_{t+1} = E_t \cup \{l\}.$$

If  $A_t$  is empty, then so is  $A_{t+1}$  and  $E_{t+1}$  will be same as  $E_t$ .

Note that during the evolution of the process, each clause is examined only once. Also every literal in  $\bigcup_t (A_t \cup E_t)$  can be reached from  $x$  via a directed path (consisting of strongly distinct literals).

For a subset  $S$  of literals, we can partition it as  $S = S^+ \cup S^-$  where  $S^+$  (resp.  $S^-$ ) is the set of all positive (resp. negative) literals of  $S$ . Let  $u_t, a_t^+, a_t^-$  be the shorthand for  $|U_t|, |A_t^+|$  and  $|A_t^-|$  where  $|\cdot|$  means the size of a set. Set  $a_t := |A_t| = a_t^+ + a_t^-$ .

**Distribution of the process.** The stochastic description of the evolution of the process  $(u_t, a_t^+, a_t^-)$ ,  $0 \leq t \leq n$  for a random 2-SAT formula on  $n$  variables can be summarized in the next lemma whose proof is immediate.

**Lemma 5** Define a triangular array of independent Bernoulli random variables.

$$W_i^{(t)} \sim Ber(\alpha_0/2n); \quad X_i^{(t)}, Y_i^{(t)} \sim Ber(\alpha_1/2n); \quad Z_i^{(t)} \sim Ber(\alpha_2/2n), \quad 1 \leq i \leq n, 0 \leq t \leq n.$$

Let  $A_t \neq \emptyset$ . Given  $H(t)$ , the history up to  $t$  and that the current literal at time  $t$  is positive, we have

$$u_t - u_{t+1} \stackrel{d}{=} \sum_{i=1}^{u_t} \{(W_i^{(t)} + X_i^{(t)}) \wedge 1\} = Bin(u_t, \alpha_0/2n + \alpha_1/2n - \alpha_0\alpha_1/4n^2)$$

$$a_{t+1}^+ - a_t^+ \stackrel{d}{=} -1 + \sum_{i=1}^{u_t} X_i^{(t)} = -1 + Bin(u_t, \alpha_1/2n), \quad a_{t+1}^- - a_t^- \stackrel{d}{=} \sum_{i=1}^{u_t} W_i^{(t)} = -1 + Bin(u_t, \alpha_0/2n).$$

Similarly, given  $H(t)$  and conditional on the event that the current literal at time  $t$  is negative, we have

$$u_t - u_{t+1} \stackrel{d}{=} \sum_{i=1}^{u_t} \{(Y_i^{(t)} + Z_i^{(t)}) \wedge 1\} = Bin(u_t, \alpha_1/2n + \alpha_2/2n - \alpha_1\alpha_2/4n^2)$$

$$a_{t+1}^+ - a_t^+ \stackrel{d}{=} -1 + \sum_{i=1}^{u_t} Z_i^{(t)} = -1 + Bin(u_t, \alpha_2/2n), \quad a_{t+1}^- - a_t^- \stackrel{d}{=} \sum_{i=1}^{u_t} Y_i^{(t)} = -1 + Bin(u_t, \alpha_1/2n).$$

**Definition 2** For the rest of the paper, we fix  $T = \lceil \sqrt{n} \rceil$ . Let  $\tau := \sup\{t \leq T : u_t \geq u_0 - 2\alpha T\}$  where  $\alpha = \max(\alpha_0, \alpha_1, \alpha_2)$ . In words,  $\tau$  is the last time before  $T$  such that the decrease in the number of unexposed variables is at most  $2\alpha T$ .

We define a round of the exploration process as follows. We fix a subset  $S$  of variables of size  $N \geq (1 - \delta/2)n$  for some small  $\delta > 0$  such that  $(1 - \delta)\rho > 1$  and a starting literal, say  $x \in S$ . We first run the exploration process from  $x$  on the implication digraph restricted to literals from  $S$  up to time  $\tau$ . If  $\tau < T$ , we stop. Otherwise, we delete all the variables in  $var((E_T \cup A_T) \setminus var(x))$  from  $S$  to get a new set of variables  $S' \subseteq S$ . By the definition of  $\tau$ ,  $|S'| \geq N - 2\alpha T$ . Now we again run another independent exploration process starting from  $\bar{x}$  up to time  $\tau$  but on the digraph restricted to literals from  $S'$ . We again have the stopping rule of  $\tau < T$ .

**Lemma 6** *The (random) set of clauses examined during the evolution of an exploration process up to time  $T$  is disjoint from the set  $\{(\bar{u} \vee \bar{v}) : u, v \in A_T\}$ . Further, the clauses examined during the evolution of the second exploration process of a round are distinct from the clauses in the set  $\{(\bar{u} \vee \bar{v}) : u, v \in A_T\} \cup \mathcal{D}$  where  $A_T$  (resp.  $\mathcal{D}$ ) is the active set at time  $T$  (resp. the set of clauses examined during the evolution) of the first exploration process.*

**Proof:** The first statement of the lemma follows from the easy observation that if  $u \in A_T$ , then, from the very definition of the exploration process, the clause  $(\bar{u} \vee w)$  is not examined up to time  $T$  for all literals  $w$  such that  $\bar{w} \notin E_T$ .

For the second statement, note that any problematic clause should include literal  $x$  or  $\bar{x}$  ( $x$  = starting vertex). Now all the clauses involving  $\text{var}(x)$  which are examined during the first exploration process of a round must have the form  $(\bar{x} \vee y)$  where literal  $y$  is such that  $\text{var}(y) \neq \text{var}(x)$ . But the clauses involving  $\text{var}(x)$  scanned by second exploration process of the round are all of the form  $(x \vee y)$  where literal  $y$  is such that  $\text{var}(y) \neq \text{var}(x)$  and there can be no clause from the set  $\{(\bar{u} \vee \bar{v}) : u, v \in A_T\}$  which contains either  $x$  or  $\bar{x}$ .  $\square$

**Corollary 1** *Given whether each of the clauses in  $\{(\bar{u} \vee \bar{v}) : u, v \in A_T\} \cup \mathcal{D}$  is present in  $\mathcal{C}$  or not, the distribution of the evolution the second exploration process only depends on number of the variables with which the second process starts.*

#### 4.1 Proof of the Theorem part (b) when $\alpha_i$ 's are all equal

Before tackling the general situation we pause for a moment to give a quick sketch, after [Ver99], of unsatisfiability part of the phase transition for the standard 2-SAT model. This will serve as a prelude to the proof for the general case.

Let  $\alpha = \alpha_0 = \alpha_1 = \alpha_2 > 1$ . Then  $\rho = \alpha$ . In this special case, we slightly modify our exploration process by demanding that we will always choose the current literal from the set of active literals uniformly. Thus at each time  $t \geq 1$ , given its size  $a_t$ ,  $A_t$  is uniformly random over all the literals except  $x$  and  $\bar{x}$ , the starting vertex and its complement.

Since the probabilities for the clauses to be present are all equal, the distribution of the exploration process  $(u_t, a_t)$  simplifies. Given  $H(t)$ , the history up to time  $t$  and  $a_t > 0$ ,

$$u_t - u_{t+1} \stackrel{d}{=} \text{Bin}(u_t, 2p_n - p_n^2), \quad a_{t+1} - a_t \stackrel{d}{=} -1 + \text{Bin}(2u_t, p_n), \text{ where } p_n = \alpha/2n. \quad (12)$$

Note that each of the random variables  $(u_t - u_{t+1})$  is stochastically dominated by  $\text{Bin}(n, 2p_n)$  which has mean  $\alpha$ . Thus using concentration of the Binomial distribution it is easy to see that for time  $T = [\sqrt{n}]$ , the event  $\{\tau < T\} = \{\sum_{t=1}^{T-1} (u_{t-1} - u_t) > 2\alpha T\}$  occurs with probability at most  $A \exp(-cT)$  where  $c > 0$ . See Lemma 7 in Section 4.2 for a proof of a more general fact.

Let  $\delta > 0$  be as given in Definition 2. If  $u_0 \geq (1 - \delta/2)n$ , then  $\{\tau = T\} \subseteq \{u_T \geq (1 - \delta)n\}$ . When  $\tau = T$ , the process  $\{a_t, t \geq 1\}$  behaves like a random walk with positive drift on nonnegative integers with 0 as the absorbing state and hence

$$\exists C > 0 \text{ such that } \mathbb{P}(a_T \geq CT, \tau = T) \geq \zeta \text{ for some constant } \zeta > 0, \text{ independent of } n. \quad (13)$$

If both  $u$  and  $\bar{u}$  are in  $A_T$  for some literal  $u$ , we have a directed path from the starting vertex to its complement using the literals in  $E_T \cup A_T$ . Else, each pair of literals  $u, v \in A_T$  are strongly distinct. There are edges  $u \rightarrow \bar{v}$  and  $v \rightarrow \bar{u}$  in the digraph if the clause  $(\bar{u} \vee \bar{v})$  is present in the formula. If at least one of these  $\binom{CT}{2}$  many clauses is present, then we again have a directed path from the starting vertex to its

complement using the literals in  $E_T \cup A_T$ . Let  $D$  be the event that there exists a directed path from the starting vertex of the exploration process to its complement in  $E_T \cup A_T$ . Therefore, by Lemma 6,

$$\mathbb{P}(D \mid a_T \geq CT, \tau = T) \geq 1 - (1 - \alpha/2n)^{\binom{CT}{2}} \geq p, \text{ where } p > 0 \text{ is a constant, independent of } n.$$

which implies that

$$\mathbb{P}(D, \tau = T) \geq \mathbb{P}(D \mid a_T \geq CT, \tau = T) \mathbb{P}(a_T \geq CT, \tau = T) \geq p\xi > 0.$$

Now, by Corollary 1, we can say that after a round, the probability that there is no termination and there exists a contradictory cycle in the variables visited during the round is at least  $p^2\zeta^2$ .

We continue with another round of exploration process in the deleted graph containing only unvisited variables. We repeat this process until a round stops due to the stopping rule. If each of the successive rounds does not terminate, we have  $\Theta(\sqrt{n})$  rounds of the exploration processes before the event  $\{u_t < (1 - \delta/2)n\}$  occurs. It is easy to see the clauses examined in different rounds of exploration process are all distinct and hence the rounds are independent.

Thus, the probability that we get no contradictory cycle in all the rounds is at most

$$\begin{aligned} & \mathbb{P}(\text{no contradictory cycle and no round stops}) + \mathbb{P}(\text{one of the rounds stops}) \\ & \leq (1 - p^2\zeta^2)^{\Theta(\sqrt{n})} + (1 - (1 - A \exp(-cT))^{2\Theta(\sqrt{n})}) \\ & \leq \text{Const} \times \exp(-B\sqrt{n}) \quad \text{for some } B > 0. \end{aligned}$$

□

**Remark 2** Instead of taking  $T = [\sqrt{n}]$  as in the proof if we choose  $T = \Theta(n)$  suitably, then (13) yields that  $a_T \geq \Omega(n)$  with probability at least  $r$  with some  $r > 0$ . Thus for any literal  $y \neq x, \bar{x}$ , we get  $\mathbb{P}(y \in A_T) \geq p$  for some  $p > 0$  and for all  $n$  large enough. So, the probability that there is a directed path from  $x_1$  to  $x_2$  is at least  $p$ . The same holds true for directed path from  $x_2$  to  $x_1$ . These are monotonic events. So, by the FKG inequality [FKG71], they occur simultaneously with probability greater than or equal to  $p^2$ . Thus the chance that there exists a directed path from  $x_1$  to  $\bar{x}_1$  is at least  $p^2$ . Again applying FKG, we have a contradictory cycle with probability at least  $p^4$ . Now appealing to Friedgut's theorem on sharp threshold [Fri99], we can conclude that the formula is UNSAT with probability tending to one as  $n \rightarrow \infty$ .

## 4.2 Associated 2-type Branching Process

Now we are back to the general case. Given an exploration process on a subgraph of the implication digraph consisting of  $N = \Theta(n)$  many variables starting from any fixed literal, our goal is to couple it with a suitable 2-type supercritical branching process up to time  $T = [\sqrt{n}]$  on a set of high probability. Assume  $N \geq (1 - \delta/2)n$  where  $\delta > 0$  is such that  $(1 - \delta)\rho > 1$ .

Thus on the set where  $\{u_t \geq N - 2\alpha T, \forall t \leq T\}$ , for large enough  $n$ ,  $\text{Bin}(u_t, \alpha_i/2n)$  stochastically dominates  $\text{Bin}((1 - \delta)n, \alpha_i/2n)$  for all time  $t \leq T$ . Next we are going to prove that this event happens with high probability.

**Lemma 7** Let  $T, \delta$  be as above and  $\alpha = \max(\alpha_0, \alpha_1, \alpha_2)$ . Then  $\mathbb{P}(\tau < T) \leq 2 \exp(-\alpha T/2)$ . Therefore,  $\mathbb{P}(u_t \geq (1 - \delta)n \ \forall t \leq T) \geq 1 - 2 \exp(-\alpha T/2)$  for sufficiently large  $n$ .

**Proof.** Since  $u_t$  is decreasing in  $t$  and  $N - 2\alpha T \geq (1 - \delta)n$  for sufficiently large  $n$ , it is enough to prove that

$$\mathbb{P}(\tau < T) \leq 2 \exp(-\alpha T/2).$$

Note that  $u_0 = N$ . Clearly,  $u_{t-1} - u_t$  is conditionally independent of  $u_0 - u_1, u_1 - u_2, \dots, u_{t-2} - u_{t-1}$  given  $u_{t-1}$  and the type the current literal at time  $t$  and is stochastically dominated by  $\text{Bin}(2N, \alpha/2n)$  irrespective of the conditioning event. Therefore, the distribution of  $u_0 - u_T$  is stochastically dominated by  $\text{Bin}(2NT, \alpha/2n)$ . By Bernstein's inequality,

$$\mathbb{P}(\text{Bin}(2NT, \alpha/2n) \geq 2\alpha T) \leq 2 \exp(-\alpha T/2).$$

Therefore,

$$\mathbb{P}(\tau < T) = \mathbb{P}(u_T < N - 2\alpha T) = \mathbb{P}(u_0 - u_T > 2\alpha T) \leq 2 \exp(-\alpha T/2).$$

□

**Lemma 8** *There exist bounded distributions  $F_i$  with mean  $m_i, 0 \leq i \leq 2$  such that*

1. *For all sufficiently large  $n$*

$$\mathbb{P}(\text{Bin}(n(1 - \delta), \alpha_i/2n) = k) \geq \mathbb{P}(X = k | X \sim F_i) \quad \forall k \geq 1.$$

2. *If  $\mathbf{M}_0$  is the branching matrix given by*

$$\mathbf{M}_0 = \begin{bmatrix} m_1 & m_0 \\ m_2 & m_1 \end{bmatrix}, \quad (14)$$

*then  $\rho_0 > 1$  where  $\rho_0$  is the maximum eigenvalue of  $\mathbf{M}_0$ .*

**Proof.** Fix some  $\beta \in (0, 1)$  so that  $(1 - \delta)(1 - \beta)\rho > 1$ . Let  $\gamma_i = (1 - \delta)\alpha_i/2$  for  $i = 0, 1, 2$ . Find  $c$  large enough so that

$$\sum_{k=1}^c k(1 - \beta/2) \exp(-\gamma_i) \gamma_i^k / k! > (1 - \beta)\gamma_i \quad i = 0, 1, 2$$

For each  $0 \leq i \leq 2$ , let us now define a truncated (and reweighed) Poisson distribution which takes the value  $k$  with probability  $(1 - \beta/2) \exp(-\gamma_i) \gamma_i^k / k!$  for  $1 \leq k \leq c$  and 0 otherwise. Call this distribution  $F_i$ . By the choice of  $c$ , its mean  $m_i$  is greater than  $(1 - \beta)\gamma_i$ . Poissonian convergence says  $\text{Bin}(n(1 - \delta), \alpha_i/2n) \xrightarrow{L^1} \text{Poisson}(\gamma_i)$  and conclusion 1 of the lemma follows.

Also,  $\rho_0 = m_1 + \sqrt{m_0 m_2} \geq (1 - \delta)(1 - \beta)\rho > 1$ . □

**Definition 3** *Let us have a supercritical 2-type branching process, which we will call an  $F$ -branching process, with offspring distributions as follows*

$$\text{1st type} \rightarrow (\text{1st type}, \text{2nd type}) : \underbrace{(F_1, F_0)}_{\text{indep}} \quad \text{2nd type} \rightarrow (\text{1st type}, \text{2nd type}) : \underbrace{(F_2, F_1)}_{\text{indep}} \quad (15)$$

Next we define a new process  $\mathbf{X}(t) = (X_1(t), X_2(t))$  by sequentially traversing the Galton-Watson tree of the  $F$ -Branching process. We fix a suitable order among the types of nodes of the tree and moreover we always prefer to visit a node of type I to a node of type II. Then we traverse the tree sequentially and at each step we expand the tree by including all the children of the node we visit. Let us denote number of unvisited or unexplored children of type  $i$  in the tree traversed up to time  $t$  by  $X_i(t)$ .

**Lemma 9** *There exists a coupling such that  $(a_t^+, a_t^-) \geq \mathbf{X}(t)$  for all  $t \leq \tau$  and  $n$  large enough.*

**Proof.** Fix  $n$  sufficiently large. If the starting vertex of exploration process is of positive type, we initiate the branching process with one individual of type I. Similarly for the other case. We run in parallel the exploration process where the choice of the type of the current literal at time  $t$  depends on the type of the visited node at time  $t$ . It can be done because, if  $t \leq \tau$ , we can always simultaneously choose our random variables in such a way (by Lemmas 5 and 8) that for every step, the number of active literals generated of each type is no less than the number of unvisited nodes of the corresponding type in the tree grown up to that step. If  $\tau < t \leq T$  or if we have no unvisited child left in the tree then we choose the current literal from the active set in some fixed predetermined procedure.  $\square$

Next we are going to find a lower bound on the total number of unvisited children after  $T$  steps of the above process.

**Lemma 10** Suppose  $\mathbf{X}(t)$  be as in Definition 3 with  $\mathbf{X}(0) = (1, 0)$  or  $(0, 1)$ . Then  $\exists C > 0, \eta > 0$  such that  $\mathbb{P}(X_1(T) + X_2(T) \geq CT) \geq \eta$ .

**Proof.** Though a proof of the above lemma can be found implicitly in [KS67], we here present it for sake of completeness. Recall that the F-branching process is supercritical as  $\rho_0$ , the maximum eigenvalue of  $\mathbf{M}_0$ , is strictly greater than 1.

If we assume  $\alpha_1 > 0$ , trivially, this process is positive regular and non-singular. Thus, by a well-known result (see [Har02]) on supercritical multitype branching process, its extinction probability is given by  $\mathbf{0} \leq \mathbf{q} = (q_1, q_2) < \mathbf{1}$  where  $q_i$  is the probability that the process becomes extinct starting with one object of type  $i$ .

Note if  $\alpha_1 = 0$ , we no longer have the positive regularity. In that case, though the above theorem can not be directly applied, we can argue as follows to get the same conclusion. If the process starts with only one individual of type  $i$ , the corresponding branching process can be viewed as a single type supercritical branching process (made of the individuals of type  $i$  only) if we observe the process only at the even number of steps. So, the probability that it eventually dies out, which is nothing but  $q_i$ , is strictly less than one.

Let us denote  $e_1 = (1, 0), e_2 = (0, 1)$ . Instead of looking at  $\mathbf{X}(t)$  which has  $(0, 0)$  as an absorbing state, we will consider a new chain  $\hat{\mathbf{X}}(t)$  starting from  $\hat{\mathbf{X}}(0) = \mathbf{X}(0)$  which is supported on entire  $\mathbb{Z}^2$ . Given  $\hat{\mathbf{X}}(0), \hat{\mathbf{X}}(1), \dots, \hat{\mathbf{X}}(t)$ , define

$$\hat{\mathbf{X}}(t+1) \stackrel{d}{=} \begin{cases} \hat{\mathbf{X}}(t) - e_1 + (F_1, F_0) & \text{if } \hat{X}_1(t) > 0 \\ \hat{\mathbf{X}}(t) - e_2 + (F_2, F_1) & \text{o.w.} \end{cases} \quad (16)$$

We can couple  $\mathbf{X}(t)$  and  $\hat{\mathbf{X}}(t)$  together so that  $\hat{\mathbf{X}}(t) = \mathbf{X}(t)$  until  $\mathbf{X}(t)$  reaches  $(0, 0)$ .

Let  $(a, b)$  be a normalized eigenvector of  $\mathbf{M}_0$  corresponding to eigenvalue  $\rho_0$  so that  $a^2 + b^2 = 1$ . Since  $\alpha_0, \alpha_2 > 0$  we have both  $a > 0$  and  $b > 0$ . Let  $Z(t) := aX_1(t) + bX_2(t)$  and  $\hat{Z}(t) := a\hat{X}_1(t) + b\hat{X}_2(t)$ . Let  $\mathcal{F}_t := \sigma(\hat{\mathbf{X}}(0), \hat{\mathbf{X}}(1), \dots, \hat{\mathbf{X}}(t))$  and  $\Delta\hat{Z}(t) := \hat{Z}(t+1) - \hat{Z}(t)$ . Since  $F_i$ 's are bounded, so are  $\Delta\hat{Z}(t)$ 's. Then

$$\mathbb{E}(\Delta\hat{Z}(t)|\mathcal{F}_t) = \begin{cases} (\rho_0 - 1)a & \text{if } \hat{X}_1(t) > 0 \\ (\rho_0 - 1)b & \text{o.w.} \end{cases} \geq \mu := (\rho_0 - 1) \min(a, b) > 0 \quad (17)$$

Now we have,

$$\begin{aligned} \mathbb{P}(\hat{Z}(T) \leq \mu T/2) &\leq \mathbb{P}\left(\hat{Z}(T) - \sum_{i=0}^{T-1} \mathbb{E}(\Delta\hat{Z}(i)|\mathcal{F}_i) \leq -\mu T/2\right) \\ &\leq \frac{\sum_{i=0}^{T-1} \mathbb{E}(\Delta\hat{Z}(i) - \mathbb{E}(\Delta\hat{Z}(i)|\mathcal{F}_i))^2}{\mu^2 T^2/4} = O(T^{-1}) \end{aligned}$$

In the last line we use orthogonality of the increments, boundedness of  $\Delta \hat{Z}(t)$  and Chebyshev inequality. Therefore we can conclude,

$$\begin{aligned}
\mathbb{P}(X_1(T) + X_2(T) \geq \mu T/2 | \mathbf{X}(0) = e_i) &\geq \mathbb{P}(Z(T) \geq \mu T/2 | \mathbf{X}(0) = e_i) \\
&\geq \mathbb{P}(\hat{Z}(T) \geq \mu T/2, \mathbf{X}(t) \neq 0 \ \forall 0 \leq t \leq T | \mathbf{X}(0) = e_i) \\
&\geq \mathbb{P}(\mathbf{X}(t) \neq 0 \ \forall t \geq 0 | \mathbf{X}(0) = e_i) - \mathbb{P}(\hat{Z}(T) < \mu T/2 | \mathbf{X}(0) = e_i) \\
&\geq (1 - q_i) - O(T^{-1}).
\end{aligned}$$

The second inequality uses the fact that  $Z(t) = \hat{Z}(t)$  until  $\mathbf{X}(t)$  reaches  $(0, 0)$ . □

## 5 Proof of the Theorem 1 part (b)

**Lemma 11** *In one round of exploration process on a subgraph involving  $N \geq (1 - \delta/2)n$  many variables the probability that (i) there is no termination due to the stopping rule and (ii) there exists a contradictory cycle using the variables visited through the round is at least  $p^2\zeta^2$  for some  $p, \zeta > 0$ .*

**Proof.** From Lemma 9 and 10, we get

$$\begin{aligned}
\mathbb{P}(a_T \geq CT, \tau = T) &\geq \mathbb{P}(X_1(T) + X_2(T) \geq CT, \tau = T) \\
&\geq \mathbb{P}(X_1(T) + X_2(T) \geq CT) - \mathbb{P}(\tau < T) \\
&\geq \eta - 2 \exp(-\alpha T/2) \geq \zeta > 0.
\end{aligned}$$

If both  $u$  and  $\bar{u} \in A_T$  for some literal  $u$ , we have a directed path from the starting vertex to its complement using the literals in  $E_T \cup A_T$ . Otherwise, for each pair of literals  $u, v \in A_T$ , which are strongly distinct, there are edges  $u \rightarrow \bar{v}$  and  $u \rightarrow \bar{v}$  in the digraph if the clause  $(\bar{u} \vee \bar{v})$  is present in the formula. If at least one of these  $\binom{CT}{2}$  many clauses is present, then we again have a directed path from starting vertex to its complement using the literals in  $E_T \cup A_T$ .

**Case I:  $\alpha_1 > 0$ .** Let  $\alpha_{\min} = \min(\alpha_0, \alpha_1, \alpha_2) > 0$ . Let  $D$  be the event that  $\exists$  a directed path from the starting vertex of the exploration process to its complement in  $E_T \cup A_T$ . Then, by lemma 6,

$$\mathbb{P}(D | a_T \geq CT \text{ and } \tau = T) \geq 1 - (1 - \alpha_{\min}/2n)^{\binom{CT}{2}} \geq p > 0.$$

**Case II:  $\alpha_1 = 0$ .** Now  $\alpha_{\min} = 0$  and we can not make the above statement. But then, instead of looking at all  $(\bar{u} \vee \bar{v})$  clauses where  $u, v$  are strongly distinct clauses belonging to  $A_T$ , we only consider such clauses where  $u$  and  $v$  have same parity. Since we have at least  $2\binom{CT/2}{2}$  many clauses of that type, we have, similarly to case I,

$$\mathbb{P}(D | a_T \geq CT \text{ and } \tau = T) \geq 1 - (1 - \alpha'/2n)^{2\binom{CT/2}{2}} \geq p > 0.$$

where  $\alpha' = \min(\alpha_0, \alpha_2) > 0$ .

Therefore,

$$\mathbb{P}(D, \tau = T) \geq \mathbb{P}(D | a_T \geq CT \text{ and } \tau = T) \mathbb{P}(a_T \geq CT \text{ and } \tau = T) \geq p\zeta.$$

The lemma is now immediate from Corollary 1. □

**Remark 3** From the proof of the above lemma, we have seen that for large  $n$  with probability at least  $r$  for some  $r > 0$  we have a directed path in the implication digraph from any literal  $u$  to its complement  $\bar{u}$ . Invoking FKG, we can say that we can find a contradictory cycle with probability at least  $r^2 > 0$ . But note that we do not have a ready-made theorem like Friedgut's sharp threshold result for generalized 2-SAT model. Though we believe that tweaking the lemma 4 of [HM06] may help, we will not pursue that. Instead we take a different route to bypass the problem.

**Proof of the Theorem 1 part (b)** We now show how to bootstrap this positive probability event to an event with high probability.

Initially we run a round of exploration process on the entire set of variables starting from  $x_1$ . If the process does not terminate after the first round, out of at least  $n - 4\alpha T - 1$  many unvisited variables, we pick up an arbitrary one and run another round of exploration process in the deleted graph starting from it. We repeat this process  $\delta\sqrt{n}/9\alpha \leq \delta n/2(4\alpha T + 1)$  many times, provided that we do not have to stop before, each time discarding previously visited variables to achieve independence among the different rounds. We thus ensure that in each run of exploration process, we have at least  $(1 - \delta/2)n$  many variables to start with.

We conclude that the probability that we get no contradictory cycle in all the rounds is at most

$$\begin{aligned} & \mathbb{P}(\text{no contradictory cycle and no round stops}) + \mathbb{P}(\text{one of the rounds stops}) \\ & \leq (1 - p^2\zeta^2)^{\delta\sqrt{n}/9\alpha} + (1 - (1 - 2\exp(-\alpha T/2)))^{2\delta\sqrt{n}/9\alpha} \\ & \leq \text{Const} \times \exp(-B\sqrt{n}) \quad \text{for some } B > 0. \end{aligned}$$

This concludes the proof. □

## References

- [Ach00] Dimitris Achlioptas. Setting 2 variables at a time yields a new lower bound for random 3-SAT (extended abstract). In *Proceedings of the Thirty-Second Annual ACM Symposium on Theory of Computing*, pages 28–37 (electronic), New York, 2000. ACM.
- [Ald01] D. Aldous. The  $\zeta(2)$  limit in random assignment problem. *Random Structures and Algorithms*, 18:381–418, 2001.
- [AP04] Dimitris Achlioptas and Yuval Peres. The threshold for random  $k$ -SAT is  $2^k \log 2 - O(k)$ . *J. Amer. Math. Soc.*, 17(4):947–973 (electronic), 2004.
- [APT82] Bengt Aspvall, Michael F. Plass, and Robert Endre Tarjan. Erratum: “A linear-time algorithm for testing the truth of certain quantified Boolean formulas” [Inform. Process. Lett. **8** (1979), no. 3, 121–123; MR 80b:68050]. *Inform. Process. Lett.*, 14(4):195, 1982.
- [ART06] Dimitris Achlioptas and Federico Ricci-Tersenghi. On the solution-space geometry of random constraint satisfaction problems. In *STOC’06: Proceedings of the 38th Annual ACM Symposium on Theory of Computing*, pages 130–139, New York, 2006. ACM.
- [BBC<sup>+</sup>01] Béla Bollobás, Christian Borgs, Jennifer T. Chayes, Jeong Han Kim, and David B. Wilson. The scaling window of the 2-SAT transition. *Random Structures Algorithms*, 18(3):201–256, 2001.
- [Bol01] Béla Bollobás. *Random graphs*, volume 73 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, second edition, 2001.

[CFS02] C. Cooper, A. Frieze, and G.B. Sorkin. A note on random 2-SAT with prescribed literal degrees. In *Proceedings of the thirteenth annual ACM-SIAM symposium on Discrete algorithms*, pages 316–320. Society for Industrial and Applied Mathematics Philadelphia, PA, USA, 2002.

[Coo71] S. A. Cook. The complexity of theorem-proving procedures. In *Proc. 3rd ACM Symposium on Theory of computing*, pages 151–158, 1971.

[CR92] V. Chvátal and B. Reed. Mick gets some (the odds are on its side). In *33rd FOCS*, pages 620–627, London, 1992. IEEE press.

[FdLV92] W. Fernandez de la Vega. On random 2-sat. *Unpublished manuscript*, 1992.

[FKG71] C. M. Fortuin, P. W. Kasteleyn, and J. Ginibre. Correlation inequalities on some partially ordered sets. *Comm. Math. Phys.*, 22:89–103, 1971.

[Fri99] Ehud Friedgut. Sharp thresholds of graph properties, and the  $k$ -sat problem. *J. Amer. Math. Soc.*, 12(4):1017–1054, 1999. With an appendix by Jean Bourgain.

[Goe92] Andreas Goerdt. A threshold for unsatisfiability. In *Mathematical foundations of computer science 1992 (Prague, 1992)*, volume 629 of *Lecture Notes in Comput. Sci.*, pages 264–274. Springer, Berlin, 1992.

[Goe96] Andreas Goerdt. A threshold for unsatisfiability. *J. Comput. System Sci.*, 53(3):469–486, 1996. 1994 ACM Symposium on Parallel Algorithms and Architectures (Cape May, NJ, 1994).

[Har02] Theodore E. Harris. *The theory of branching processes*. Dover Phoenix Editions. Dover Publications Inc., Mineola, NY, 2002. Corrected reprint of the 1963 original [Springer, Berlin; MR0163361 (29 #664)].

[HM06] H. Hatami and M. Molloy. Sharp thresholds for constraint satisfaction problems and homomorphisms. <http://arxiv.org/abs/math/0612391>, 2006.

[JŁR00] Svante Janson, Tomasz Łuczak, and Andrzej Rucinski. *Random graphs*. Wiley-Interscience Series in Discrete Mathematics and Optimization. Wiley-Interscience, New York, 2000.

[KS67] H. Kesten and B. P. Stigum. Limit theorems for decomposable multi-dimensional Galton-Watson processes. *J. Math. Anal. Appl.*, 17:309–338, 1967.

[MMW07] Elitza Maneva, Elchanan Mossel, and Martin J. Wainwright. A new look at survey propagation and its generalizations. *J. ACM*, 54(4):Art. 17, 41 pp. (electronic), 2007.

[MMZ06] Stephan Mertens, Marc Mézard, and Riccardo Zecchina. Threshold values of random  $K$ -SAT from the cavity method. *Random Structures Algorithms*, 28(3):340–373, 2006.

[MWW08] E. Mossel, D. Weitz, and N. Wormald. On the hardness of sampling independent sets beyond the tree threshold. To Appear in Prob. Theory Related. Fields, 2008.

[PZ02] M. Mézard G. Parisi and R. Zecchina. Analytic and algorithmic solution of random satisfiability problems. *Science*, 297, 812, 2002. (Scienceexpress published on-line 27-June-2002; 10.1126/science.1073287).

[Ver99] Y. Verhoeven. Random 2-SAT and unsatisfiability. *Information Processing Letters*, 72:119–123, 1999.

[Wei06] Dror Weitz. Counting independent sets up to the tree threshold. In *STOC'06: Proceedings of the 38th Annual ACM Symposium on Theory of Computing*, pages 140–149, New York, 2006. ACM.